

Przestrzeń, fluktuacje i przepływ informacji w matematycznym modelowaniu ścieżek sygnałowych

Paulina Szymańska-Rożek

autoreferat rozprawy doktorskiej

Tematem rozprawy są trzy aspekty modelowania matematycznego ścieżek sygnałowych - przestrzenność, stochastyczność i przepływ informacji. Każdy z badanych aspektów został pokazany w trzech różnych modelach biologii teoretycznej: odwracalnych cyklach fosorylacji-defosorylacji, przełączniku genetycznym i modelu aktywacji genu przez inny gen, ze sprzężeniem zwrotnym i bez.

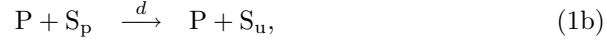
Modele i opis matematyczny

W rozprawie doktorskiej przedstawiam wybrane zagadnienia z dziedziny modelowania matematycznego ścieżek sygnałowych na przykładzie modeli zaczerpniętych z biochemii, biologii teoretycznej i biofizyki. Ich wspólnym matematycznym trzonem są Łańcuchy Markowa z czasem ciągłym, ponadto zostały wykorzystane narzędzia matematyczne z zakresu równań różniczkowych zwyczajnych, procesów stochastycznych i teorii informacji.

Wyniki są podzielone na trzy części, zgodnie z tytułem rozprawy. Rozdział “Space” traktuje o aspekcie przestrzennym na przykładzie cyklu fosorylacji-defosorylacji. Rozdział “Noise” ukazuje możliwości wykorzystania stochastyczności w kontrolowaniu populacji komórek z przełącznikiem genetycznym. W rozdziale “Information” optymalizuję informację wzajemną w prostym układzie biologicznym, w którym jedna wielkość biologiczna (np. koncentracja cukru, bądź aktywność genu) wpływa na inną wielkość (np. koncentrację enzymu metabolizującego cukier lub aktywność innego genu), ze sprzężeniem zwrotnym lub bez.

Przestrzenność

Przestrzenny aspekt modelowania ścieżek sygnałowych został przedstawiony na przykładzie modelu odwracalnego cyklu reakcji fosforylacji-defosforylacji. Substraty są fosforylowane i defosforylowane przez kinazy i fosfatazy w następujący sposób:



gdzie przez S_u i S_p oznaczone są substraty nieufosforylowane i ufosforylowane, K oznacza kinazy, a P – fosfatazy. Litery nad strzałkami oznaczają mikroskopowe stałe szybkości reakcji – fosforylacji, c i defosforylacji, d .

Rozważamy też model, w którym defosforylacja przebiega następująco:



a fosforylacja przebiega jak poprzednio.

Głównym naszym zadaniem było znalezienie *efektywnych makroskopowych statycznych szybkości reakcji*: c_{eff} – fosforylacji i d_{eff} – defosforylacji, które spełniałyby następujący układ równań różniczkowych:

$$\frac{d}{dt} \rho_{S_u} = -c_{\text{eff}} \rho_K \rho_{S_u} + d_{\text{eff}} \rho_P \rho_{S_p}, \quad (3)$$

$$\frac{d}{dt} \rho_{S_p} = c_{\text{eff}} \rho_K \rho_{S_u} - d_{\text{eff}} \rho_P \rho_{S_p}. \quad (4)$$

Gdy rozważa się przestrzeń w modelach matematycznych zwykle pierwszym narzędziem, które przychodzi do głowy, są równania różniczkowe cząstkowe. Tymczasem my wykorzystujemy podejście probabilistyczne – rozważamy Łańcuch Markowa, którego stany zdefiniowane są przez położenie cząsteczek (enzymów, ufosforylowanych i zdefosforylowanych substratów, ewentualnie innych niereagujących cząsteczek) w przestrzeni. Rzeczona przestrzeń jest skończoną, dwuwymiarową siatką trójkątną (tzn. każde miejsce sąsiaduje z sześcioma innymi) bez brzegu. Szybkości przejść między stanami w Łańcuchu Markowa odpowiadają stałym szybkościom reakcji, c i d (gdy odpowiednia para substrat-enzym znajduje się w sąsiadujących polach) i szybkości poruszania się cząsteczek w przestrzeni, m . Używamy w dużej mierze numerycznych symulacji (wykorzystujących właśnie algorytm kinetic Monte Carlo), ale prowadzimy też analityczne rozważania w przypadkach, gdy $m = 0$ i $m = \infty$.

W skrócie chcemy wyliczyć efektywne makroskopowe stałe szybkości reakcji zdefiniowane w następujący sposób:

$$c_{\text{eff}} = \frac{n}{\rho_{\text{S}_u} \rho_{\text{K}} V \Delta t}, \quad (5a)$$

$$d_{\text{eff}} = \frac{n}{\rho_{\text{S}_p} \rho_{\text{P}} V \Delta t} \quad (5b)$$

gdzie, n to liczba reakcji, które zaszły w przedziale czasu Δt , a V to liczba miejsc na sieci. Gęstości kinaz, fosfataz i substratów są oznaczone przez ρ z odpowiednim znacznikiem: ρ_{K} , ρ_{P} , ρ_{S_u} i ρ_{S_p} . Ponadto wyznaczaliśmy stacjonarne frakcje ufosforylowanych i zdefosforylowanych substratów: $\rho_{\text{S}_p}/\rho_{\text{S}}$ i $\rho_{\text{S}_u}/\rho_{\text{S}}$.

Wspomniane czysto analityczne podejście możemy zastosować w przypadku nieskończonej dyfuzji, tzn., kiedy $m = \infty$. Wówczas efektywne stałe szybkości reakcji wynoszą $c_{\text{eff}}^{\infty} = 6c$ i $d_{\text{eff}}^{\infty} = 6d$. Drugim przypadkiem, dla którego możemy zastosować czysto analityczne podejście, jest przypadek zerowej dyfuzji, tzn. gdy $m = 0$. Po trochę dłuższych rozważaniach kombinatorycznych znaleźliśmy wyrażenia również dla tego przypadku. Potwierdziliśmy następnie ich zgodność z symulacjami numerycznymi. Najbardziej “rzeczywisty” przypadek, kiedy dyfuzja jest skończona i niezerowa jest analitycznie niezwykle trudnym problemem. W ogólnym równaniu na dynamikę modelu przyjęliśmy, że efektywne makroskopowe stałe reakcji składają się z “członu zerowej dyfuzji”, c_{eff}^0 i d_{eff}^0 , oraz członu liniowo proporcjonalnego do współczynnika dyfuzji, $\lambda \cdot m$. Wówczas równanie różniczkowe na ewolucję frakcji ufosforylowanych substratów wygląda następująco:

$$\frac{d}{dt} \rho_{\text{S}_p} = (\lambda m + c_{\text{eff}}^0) \rho_{\text{K}} \rho_{\text{S}_u} - (\lambda m + d_{\text{eff}}^0) \rho_{\text{P}} \rho_{\text{S}_p}, \quad (6)$$

gdzie współczynnik λ musieliśmy wyznaczać na podstawie analizy numerycznej, i to tylko dla przypadku symetrycznego gdy $c = d$ i $\rho_{\text{K}} = \rho_{\text{P}}$. Trudności wynikają z faktu, że nawet dla tego symetrycznego przypadku, szczególnie gdy gęstości enzymów są małe, rozmieszczenie przestrzenne cząsteczek nie jest jednorodne i ufosforylowany substrat będzie raczej w pobliżu kinazy, a zdefosforylowany w pobliżu fosfatazy. Co za tym idzie, mimo iż całkowite prawdopodobieństwo, że substrat jest ufosforylowany wynosi 0.5, efektywne stałe szybkości reakcji są obniżone, bo enzymy stykają się głównie z substratami, z którymi już weszły w reakcję.

W najbardziej ogólnym przypadku efektywne stałe szybkości reakcji zależą nietrywialnie od obu mikroskopowych stałych szybkości reakcji, współczynnika dyfuzji i gęstości obu enzymów. Wykazaliśmy, że maleją one wraz z malejącą dyfuzją i zależność ta jest wyraźniejsza dla enzymu, którego jest mniej. Co za tym idzie stacjonarne frakcje ufosforylowanych i zdefosforylowanych substratów rosną lub maleją wraz z dyfuzją, w zależności od koncentracji enzymów. Ponadto zbadaliśmy wpływ innych czynników na efektywne stałe szybkości reakcji:

- wielkości planszy (liczby pól),

- zatłoczenia molekularnego (dodania do planszy poruszających się, ale nie-reagujących cząsteczek),
- tworzenia się bardziej trwałych kompleksów substrat-enzym, zarówno przed, jak i po reakcji.

Czując niedosyt spowodowany niemożnością analitycznego zbadania przypadku skończonej, niezerowej dyfuzji, rozważyliśmy wreszcie wariant modelu, w którym enzym i substrat muszą znaleźć się w tym samym polu, aby zaszła reakcja. Dzięki tej drobnej zmianie mogliśmy przeprowadzić rozumowanie o średnim czasie powrotu do danego pola (mean first passage time) i otrzymać zaskakująco dobre przybliżenia dla symulacji numerycznych.

Fluktuacje

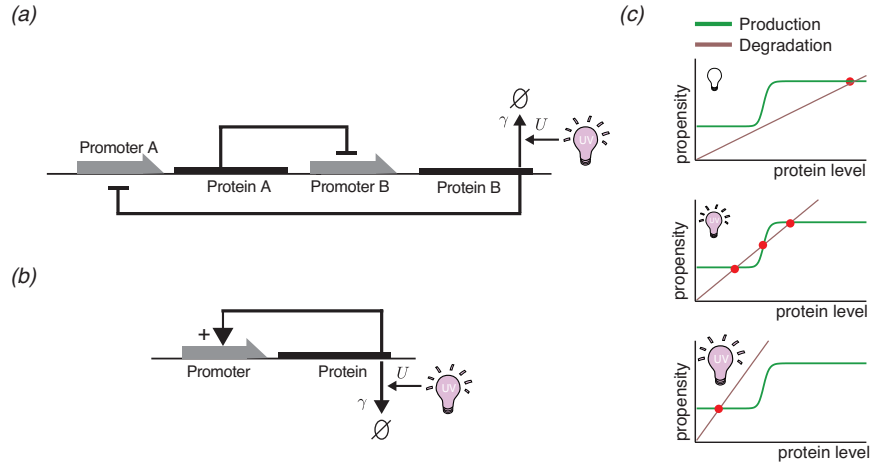
W tym rozdziale prezentuję wyniki badań rozpoczętych w ramach projektu naukowego na szkole letniej *Quantitative Biology*, która odbyła się w lipcu 2014 roku w Albuquerque (New Mexico, USA). Postawione tam pytanie badawcze brzmiało: “czy można wykorzystać fluktuacje (szum) do wyróżnienia z populacji identycznych komórek jednej, stosując jednakową strategię do całej populacji?”. Strategią miało być naświetlanie promieniowaniem UV, które wzmacnia degradację cząsteczek białka. W rozprawie zdecydowałam się użyć terminu “noise”, gdyż funkcjonuje on w anglojęzycznej literaturze naukowej. Po polsku jego dokładne tłumaczenie, “szum”, brzmi dość kolokwialnie, (mimo, że jest używane w żargonie naukowym). Ale przede wszystkim używa się go głównie w kontekście niepewności wynikającej z niedokładności narzędzi pomiarowych, a nie niepewności “wrodzonej”. Dlatego też zdecydowałam się na termin “fluktuacje”, który najlepiej oddaje kontekst angielskiego “noise”.

Identyczne komórki tworzące populację miały być rządzone przez ten sam mechanizm genetyczny. Aby można było wykorzystać fluktuacje do wyróżnienia jednej spośród nich trzeba było wybrać taki schemat regulacji, który pozwala komórce osiągnąć dwa wyraźnie różne stany - np. wysokiej i niskiej syntezy białka. Przykładem biologicznym takiego mechanizmu jest przełącznik genetyczny, w którym dwa geny nawzajem hamują swoją aktywność (blokując syntezę kodowanych białek), przedstawiony na panelu (a) rysunku 1. Prostszy model, w którym komórka może przyjąć dwa różne stany, to model samoaktywującego się genu (białko kodowane przez dany gen stanowi aktywator dla tegoż genu) - schemat ten przedstawiony jest na panelu (b) rysunku 1.

Koncentracja białka x w każdej komórce opisana jest równaniem różniczkowym zwyczajnym postaci:

$$\frac{dx}{dt} = k(x) - R(x), \quad (7)$$

gdzie $k(x)$ to funkcja produkcji: $k(x) = k_0 + k_1 x^m / (x^m + \beta^m)$, a $R(x)$ to funkcja degradacji: $R(x) = (\gamma + u)x$. Przez u oznaczyliśmy naświetlanie promieniowaniem UV. Przy odpowiednio dobranych parametrach k_0 , k_1 , β , m , γ i funkcji

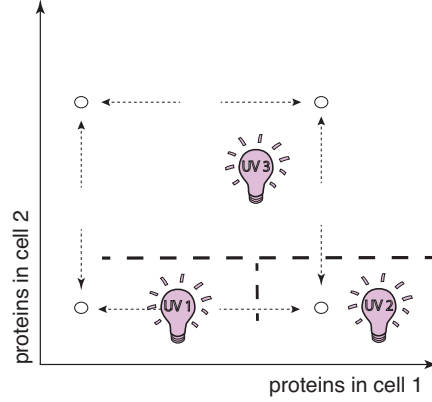


Rysunek 1: (a) Klasyczny model przełącznika genetycznego, w którym dwa rodzaje białek hamują nawzajem swoją syntezę i degradacja jednego z nich jest wzmacniona przez promieniowanie UV. (b) Model samoaktywującego się genu - uproszczona wersja przełącznika genetycznego, w którym zamiast dwóch rodzajów białek hamujących wzajemnie swoją ekspresję, jest jeden rodzaj, pobudzający swoją ekspresję. Jest to najprostszy model wykazujący bistabilność. Degradacja białka jest wzmożona przez wyższy poziom promieniowania UV. (c) Produkcja (zielona linia) i degradacja (brązowa linia) versus poziom białek, x , w modelu samoaktywującego się genu (uproszczonego przełącznika genetycznego). Punkty przecięcia $k(x) = R(x)$ wyznaczają punkty stacjonarne równania $\dot{x} = k(x) - R(x)$. Trzy panele odpowiadają przypadkom: słabego promieniowania UV (pierwszy od góry, jeden punkt stacjonarny dla wysokiej ekspresji białka), średniego promieniowania UV (środkowy rysunek, dwa stabilne punkty stacjonarne i jeden niestabilny) i wysokiego (dolny rysunek, jeden punkt stacjonarny dla niskiej ekspresji białka).

u równanie to ma trzy punkty stacjonarne, a innym wypadku jeden - dobra ilustracja różnych wariantów jest pokazana w panelu (c) rysunku 1.

Jeśli komórki są identyczne, tzn. ewolucja koncentracji białka przebiega zgodnie z powyższym równaniem, i mają ten sam warunek początkowy, to w deterministycznym ujęciu komórki te będą nierozróżnialne - w dowolnym momencie będą miały tę samą koncentrację białka. Jeśli natomiast weźmiemy pod uwagę stochastyczność, która wpisana jest we wszelkie procesy z małą liczbą elementów, liczba cząsteczek białka w komórce będzie modelowana za pomocą zmiennej losowej. W zależności od wartości przyjmowanych przez tę zmienną losową, cała populacja będzie naświetlana promieniowaniem UV z jednakowym natężeniem.

Pierwszy rozważany model składał się tylko z dwóch komórek - jedna z nich była wyróżnioną komórką, a druga “resztą populacji”. Strategię naświetlania promieniowaniem UV udało się zbudować możliwie prostą - funkcja $u := u(x)$ przyjmowała tylko trzy wartości: u_1 , u_2 i u_3 , w zależności od tego, ile było cząsteczek białka w wyróżnionej komórce i w drugiej komórce. Strategia ta została przedstawiona na rysunku 2 i opisana szczegółowo w podpisie. Wyniki w



Rysunek 2: Strategia naświetlania promieniowaniem UV dwóch komórek. Gdy obie komórki zawierają mało cząsteczek białka (lewy dolny róg), promieniowanie UV nie zostaje włączone ($u_1 = 0$). Gdy w komórce 1 jest dużo białka, a w komórce 2 mało (prawy dolny róg) włączone zostaje promieniowanie o średnim natężeniu ($u_2 = 0.75$). Gdy liczba cząsteczek białka w komórce 2 jest powyżej pewnego poziomu ($= 12$) (górna część ćwierćpłaszczyzny), włączone zostaje promieniowanie UV o silnym natężeniu ($u_3 = 1.75$). Wszystkie parametry ($k_0 = 5$, $k_1 = 50$, $\beta = 20$, $m = 4$ i $\gamma = 0.5$) są stałe i takie same dla obu komórek.

modelu dwukomórkowym były zaskakująco zadowalające - wyróżniona komórka wykazywała wyższą ekspresję białka niż druga komórka z prawdopodobieństwem 0.99. Gdy zwiększyliśmy populację do 30 komórek, dzięki tej samej strategii (natężenie promieniowania UV dobierane było na podstawie ekspresji białka w wybranej komórce i maksimum ekspresji białka w niewyróżnionych komórkach) wyróżniona komórka znajdowała się wśród 20% “najlepszych” (pod względem liczby cząsteczek białka) komórek. Ponad te dwa modele zostały rozważone inne warianty i strategie:

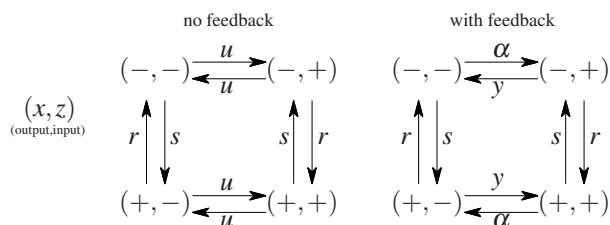
- model, w którym uwzględnione są opóźnienia czasowe w syntezie białek
- model z uwzględnionym szumem w syntezie białek (intensywności syntezy k_0 i k_1 w niewyróżnionych komórkach są mnożone przez współczynnik, który jest zmienną losową z rozkładu normalnego ze średnią 1 i odchyleniem standardowym 0.1, natomiast te współczynniki w wyróżnionej komórce wynoszą tyle, ile średnia tych współczynników w niewyróżnionych

komórkach, bądź średnia + jedno odchylenie standardowe, bądź średnia – jedno odchylenie standardowe.)

- strategię, w której dobiera się poziom natężenia promieniowania UV na podstawie innych statystyk niż maksimum ekspresji białka z niewyróżnionych komórek - ze średniej, mediany, 75-tego centyla i wyłącznie na podstawie poziomu białka w komórce wyróżnionej
- nieuproszczony model przełącznika genetycznego (z dwoma rodzajami “współzawodniczących” białek).

Przepływ informacji

Ten ciekawy aspekt modelowania matematycznego ścieżek sygnałowych interesuje nie tylko matematyków, ale w dużej mierze przyrodników - biologów, biochemików i biofizyków. Moje badania dotyczące informacji przekazywanej w układach biologicznych rozpoczęły się podczas stażu u dr Aleksandry Walczak w Ecole Normale Supérieure w Paryżu i zainspirowane były prostym pytaniem, czy i jak układy biologiczne “komunikują” się w zarówno wewnątrz, jak i z otoczeniem. Badany przeze mnie model składa się z dwóch binarnych zmiennych losowych, dla których wyliczam maksymalną informację wzajemną (ją też interpretuję jako ilość informacji przekazanej w systemie). Ponadto zakładam, że do przepływu informacji potrzebna jest energia, rodzi się więc naturalne pytanie, jak to ograniczenie energetyczne wpływa na ilość przekazanej informacji. Wreszcie stawiam często spotykane pytanie, czy sprzężenie zwrotne jest korzystne; w tym przypadku, czy zwiększa informację wzajemną rozważanych zmiennych losowych. Badany model został przedstawiony na rysunku 3.



Rysunek 3: Schemat przejść między stanami w modelu bez sprzężenia zwrotnego (lewa część) i ze sprzężeniem zwrotnym (prawa część). Zmienne są dwie i są dyskretne, wobec tego są cztery stany, szybkości przejść między nimi widnieją obok strzałek. Widoczna jest symetria między “czystymi” stanami $((-, -)$ i $(+, +)$) oraz “mieszanymi” $((-, +)$ and $(+, -)$) w obu modelach.

Głównym zadaniem jest znalezienie parametrów (u, s, r, α, y) maksymalizujących informację wzajemną pomiędzy zmienną z_0 - inputem w stanie początkowym i zmienną x_t - outputem w momencie t . Formalnie szukamy maksimum

funkcji parametrów i czasu:

$$I[X_t, Z_0] = \sum_{x_t, z_0} p(x_t, z_0) \log \frac{p(x_t, z_0)}{p(x_t)p(z_0)}, \quad (8)$$

gdzie przez $p(x_t, z_0)$ oznaczyliśmy prawdopodobieństwo, że zmienne są w stanie $(-, -)$, $(-, +)$, $(+, -)$, $(+, +)$. Jest to definicja informacji wzajemnej, która mierzy, jak wiedza o jednej zmiennej zmniejsza niepewność o drugiej zmiennej. Jest równoważna bardziej powszechnej definicji używającej *Entropii* zmiennej losowej i *Entropii warunkowej* zmiennych losowych, $I[X_t, Z_0] = S(X_t) - S(X_t|Z_0)$.

Ograniczenie energetyczne, które bierzemy pod uwagę szukając maksimum funkcji, to *szybkość produkcji entropii*. Jest to funkcja również zależna od czasu i parametrów (przejść między stanami):

$$\sigma(t) = \sum_{i,j} p_i(t) w_{ij} \log \frac{p_i(t) w_{ij}}{p_j(t) w_{ji}} \quad (9)$$

W stanie stacjonarnym wielkość ta wynosi:

$$\sigma(t) \xrightarrow{t \rightarrow \infty} \sum_{i,j} p_i^{ss} w_{ij} \log \frac{w_{ij}}{w_{ji}} = \sigma^{ss} \quad (10)$$

i to ona stanowi zapowiedziane ograniczenie energetyczne. Rozważyliśmy też inny typ ograniczenia energetycznego - *średnia dysypacja* jest całką średnią z szybkości produkcji entropii, policzoną do pewnego ustalonego momentu:

$$\Sigma^{avg}(\tau_p) = \frac{1}{\tau_p} \int_0^{\tau_p} \hat{\sigma}(\tau) d\tau. \quad (11)$$

Do tej wielkości odwołujemy się też jako do *kosztu* przesyłania informacji.

Postawiony problem jest więc dobrze zdefiniowany - dla ustalonego czasu, w którym zbieramy informację o stanie outputu (tzn. dla ustalonego τ liczymy $I[x_\tau, z_0]$) szukamy parametrów, które zwracają maksymalną informację wzajemną. Ponadto rozwiązujemy ten sam problem założywszy, że dostępna energia (stacjonarna szybkość produkcji entropii, σ^{ss}) jest ograniczona. Wreszcie liczymy koszt przesłanej maksymalnej informacji.

Rozważyliśmy cztery modele:

- S - model bez sprzężenia zwrotnego, ze stacjonarnym warunkiem początkowym
- F - model ze sprzężeniem zwrotnym, ze stacjonarnym warunkiem początkowym

- \tilde{S} - model bez sprzężenia zwrotnego, z niestacjonarnym warunkiem początkowym
- \tilde{F} - model ze sprzężeniem zwrotnym, z niestacjonarnym warunkiem początkowym

Wyniki, jakie uzyskaliśmy dla powyższych czterech modeli są zapisane w poniższej tabeli:

	I^{opt}	Koszt
S, F	$I(S) < I(F)$	$C(S) = C(F)$
\tilde{S}, \tilde{F}	$I(\tilde{S}) \leq I(\tilde{F})$	$C(\tilde{S}) > C(\tilde{F})$

Tablica 1: Porównanie czterech modeli, S , F , \tilde{S} i \tilde{F} w terminach maksymalnej informacji wzajemnej, I^{opt} , i średniej dysypacji (“kosztu”), C .

Stacjonarny warunek początkowy gwarantuje, że sprzężenie zwrotne będzie efektywniejsze w przesyłaniu informacji niż jego brak, ale koszty poniesione w obu przypadkach są takie same. Natomiast gdy system startuje z niestacjonarnego warunku początkowego, jest mniejsza różnica w przesłanej informacji (choć model ze sprzężeniem zwrotnym jest znów efektywniejszy), ale koszt jest znacząco wyższy, gdy sprzężenia zwrotnego nie ma.

Publikacje

Wyniki opisane w rozprawie doktorskiej zostały opublikowane w trzech pracach.

Rezultaty z rozdziału o przestrzenności znajdują się w [3] i [1]. W pierwszej pracy wyprowadziłam analityczne wyrażenia na efektywne makroskopowe stałe reakcji w granicy zerowej i nieskończonej dyfuzji. Dla przypadku skończonej, niezerowej dyfuzji przeprowadziłam rozumowanie oszacowujące efektywne makroskopowe stałe reakcji dla przypadku “symetrycznego”. Ponadto wykonałam symulacje stochastyczne, przygotowałam większość rysunków, które znalazły się w rozprawie doktorskiej i wreszcie napisałam sam manuskrypt, który posłużył do konstrukcji pierwszego rozdziału rozprawy. W drugiej pracy, wspólnie z pierwszym współautorem wyprowadziłam dokładniejsze oszacowania efektywnych makroskopowych stałych reakcji dla przypadku niezerowej, skończonej dyfuzji, wykorzystując teorię średniego czasu powrotu (mean first passage time). Stosowne oświadczenia od wszystkich współautorów prac są dołączone do niniejszego autoreferatu.

Badania dotyczące zagadnienia stochastyczności przedstawione zostały w [2]. W pracy przygotowałam teoretyczną podstawę do stymulacji numerycznych przeprowadzonych przez współautorów. Jestem autorką manuskryptu i przepro-

wadzonych rozumowań matematycznych. Oświadczenia współautorów dołączam do niniejszego autoreferatu.

Rozdział o przepływie informacji zawiera nieopublikowane jeszcze wyniki, które uzyskałam podczas zagranicznego stażu doktoranckiego u Aleksandry Walczak. Przeprowadziłam rozważania analityczne, dokładne rachunki komputerowe w symbolicznym programie Mathematica i numeryczną optymalizację informacji wzajemnej.

Literatura

- [1] Paweł Nałęcz-Jawecki, Paulina Szymańska, Marek Kochańczyk, Jacek Mięgisz, and Tomasz Lipniacki. Effective reaction rates for diffusion-limited reaction cycles. *J Chem Phys*, 143(21):215102, 2015.
- [2] Paulina Szymańska, Nicola Gritti, Johannes M. Keegstra, Mohammad Soltani, and Brian Munsky. Using noise to control heterogeneity of isogenic populations in homogenous environments. *Phys Biol*, 12(4):045003, 2015.
- [3] Paulina Szymańska, Marek Kochańczyk, Jacek Mięgisz, and Tomasz Lipniacki. Effective reaction rates in diffusion-limited phosphorylation-dephosphorylation cycles. *Phys Rev E*, 91:022702, Feb 2015.